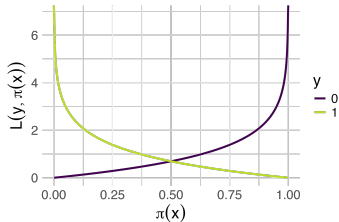
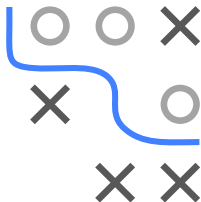


## Advanced Risk Minimization

### Proper Scoring Rules



- Scoring rules on prob. predictions
- First order condition of SRs
- Log loss and Brier are strictly proper
- L1 on probs is not
- 0/1 is proper but not strict

- Scoring rules on prob. predictions
- First order condition of SRs
- Log loss and Brier are strictly proper
- L1 on probs is not
- 0/1 is proper but not strict

- 

$$S(Q, P) = \mathbb{E}_{y \sim Q}[L(Q, P)]$$

- Simply expected loss, now we write  $P$  for  $\pi(\mathbf{x})$
- We have looked at this before
- As we have seen in the beginning of this chapter:  
Can do this unconditionally or conditionally, all the same anyway
- Minimizing the above asks then for the risk minimizer!

- SR is **proper** if true label distrib  $Q$  is among the optimal solutions, when we maximize  $S(Q, P)$  in the 2nd argument (for a given  $Q$ )

$$S(Q, Q) \leq S(Q, P) \text{ for all } P, Q$$

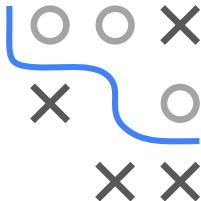
- Translation (now for conditional):  
 $\mathbb{P}(y = 1 \mid \mathbf{x} = \tilde{\mathbf{x}})$  is **among the solutions** for the RM:

$$\eta(\tilde{\mathbf{x}}) = \mathbb{P}(y = 1 \mid \mathbf{x} = \tilde{\mathbf{x}}) \in \pi^*(\tilde{\mathbf{x}})$$

- NB: We were never precise before that the RM as “argmin” is a set!
- $S$  is **strictly proper** when equality holds iff  $P = Q$
- Translation (now for conditional):  
 $\mathbb{P}(y = 1 \mid \mathbf{x} = \tilde{\mathbf{x}})$  is unique RM!

$$\pi^*(\tilde{\mathbf{x}}) = \eta(\tilde{\mathbf{x}}) = \mathbb{P}(y = 1 \mid \mathbf{x} = \tilde{\mathbf{x}})$$

- (Strictly) proper SR ensure optimization pushes solution  $\pi(\mathbf{x})$  to  $Q$



# L1 LOSS IS NOT PROPER

- Let's look at it unconditionally, so we don't always have to write  $x$
- Binary targets  $y \sim \text{Bern}(\eta)$
- For any binary prob. loss  $L$ :

$$\mathbb{E}_y[L(y, \pi)] = \eta \cdot L(1, \pi) + (1 - \eta) \cdot L(0, \pi)$$

- Let's check L1 loss  $L(y, \pi) = |y - \pi|$

$$\mathbb{E}_y[L(y, \pi)] = \eta|1 - \pi| + (1 - \eta)\pi = \eta + \pi(1 - 2\eta)$$

- Linear in  $\pi$ , but with box constraints
- For  $\eta > 0.5$ :  $\pi^* = 1$
- For  $\eta < 0.5$ :  $\pi^* = 0$
- So  $\pi^*$  usually not the same as  $\eta$
- True  $\eta$  is worse than RM  $\pi^*$ , so L1 not proper SR

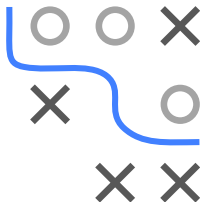


# 0/1 LOSS IS PROPER – BUT NOT STRICT

- **0/1 loss**  $L(y, \pi) = \mathbb{1}_{\{y \neq h_\pi\}}$  using hard labeler  $h_\pi = \mathbb{1}_{\{\pi \geq 0.5\}}$
- Expected loss:

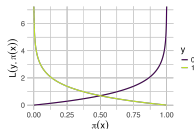
$$\begin{aligned}\mathbb{E}_y[(y, \pi)] &= \eta \cdot L(1, \pi) + (1 - \eta) \cdot L(0, \pi) \\ &= \begin{cases} \eta & \text{if } h_\pi = 0 \\ 1 - \eta & \text{if } h_\pi = 1 \end{cases}\end{aligned}$$

- So expected loss only takes 2 values
- For  $\eta \geq 0.5$ , minimal if  $h_\pi = 1$   
→ any  $\pi \in [0.5, 1]$  minimizes
- For  $\eta < 0.5$  expected loss is minimal if  $h_\pi = 0$   
→ any  $\pi \in [0, 0.5)$  minimizes
- True  $\eta$  among solutions → proper
- True  $\eta$  is not unique minimizer → not strictly proper



# DISCOVER PROPER SCORING RULES

- How define  $L$  such that  $\mathbb{E}_y[L(y, \pi)]$  is minimized at  $\pi^* = \eta$ ?
- Assume symmetry:  $L(1, \pi) = L(\pi)$  and  $L(0, \pi) = L(1 - \pi)$
- Remember that when we plotted such losses, we usually had 2 curves, symmetric at  $\pi = 0.5$ ?
- Then, with  $\eta = \mathbb{P}(y = 1)$



$$\mathbb{E}_y[L(y, \pi)] = \eta \cdot L(\pi) + (1 - \eta) \cdot L(1 - \pi)$$

- First-order condition: Set  $L'(\pi) = 0$ , and  $\pi = \eta$  at minimum

$$\eta \cdot L'(\eta) \stackrel{!}{=} (1 - \eta) \cdot L'(1 - \eta)$$

# LOG LOSS / BRIER ARE STRICTLY PROPER

- First-order condition:

$$\eta \cdot L'(\eta) \stackrel{!}{=} (1 - \eta) \cdot L'(1 - \eta)$$

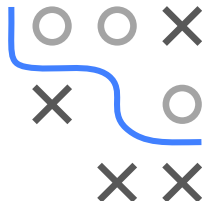
- One solution is  $L'(\eta) = -1/\eta$
- Antiderivative is  $L(\eta) = -\log(\eta)$
- Remember:  $L(1, \pi) = L(\pi)$  and  $L(0, \pi) = L(1 - \pi)$
- This is **log loss**

$$L(y, \pi) = -y \cdot \log(\pi) - (1 - y) \cdot \log(1 - \pi)$$

- Second solution is  $L'(\eta) = -2(1 - \eta)$
- Antiderivative  $L(\eta) = (1 - \eta)^2$
- So **Brier score**

$$L(y, \pi) = (y - \pi)^2$$

- Both strictly proper (check 2nd derivative for strict convexity)



# OUTLOOK

- Usually SRs are maximized, I adapted notation a bit here for us, to get a direct connection ERM
- Was easier to talk about binary classification here, but proper SRs are defined in general
- There are other proper SRs, like “generalized entropy score” or “continuous ranked probability score”
- We only scratched surface of theory
- If you want to know more: start by reading the Gneiting paper

